

PREDIKCE PŘEŽITÍ NA TITANIKU

TEORETICKÉ UKOTVENÍ

O Titaniku se říkalo, že je nepotopitelný. Když vyplul na svou první plavbu, byla na jeho palubě celá řada významných osobností, ale také mnoho obyčejných lidí putujících za svým americkým snem. Až do osudné noci 14. dubna 1912 si nikdo na palubě Titaniku nepřipouštěl nebezpečí, jakému jejich loď mřila vstříc, a které v sobě ukrývaly hory plovoucího ledu. Tehdy Titanic narazil do ledovce v Severním moři, který ho neúprosně poslal ke dnu. Klidná plavba se změnila ve zběsilý a tragický boj o záchranu života.

Potopení Titanicu patří mezi celosvětově známé tragédie. Kapacita lodi mohla převážet až 2 603 cestujících. O provoz lodě se sralo až 899 členů posádky. Při srážce s ledovcem zemřelo 1 502 z 2 224 cestujících včetně posádky. Tragédie šokovala mezinárodní společnosti a vedla k lepším bezpečnostním předpisům pro lodě. (Encyclopaedia Britannica, n.d.)

Jedním z důvodů, proč zde byla tak vysoká ztráta na životech, byl nedostatek záchranných člunů pro cestující a posádku. Některé skupiny lidí však měly větší šanci na přežití než jiné. Jací cestující tedy měly větší šanci na přežití?

STATISTICKÁ METODA – BINÁRNÍ LOGISTICKÁ REGRESE

Ke zpracování dat byla využita metoda binární logistické regrese. Metoda umožňuje odhad pravděpodobnosti nastoupení nějakého jevu (zapíšeme ho pomocí náhodné veličiny X jako $(Y=1)$ pomocí známých regresorů $X_1, \dots, X_k$, které mohou být jak spojitého, tak kategoriálního typu. Jedná se o statistický model, který popisuje chování dichotomické závisle proměnné.

DATOVÝ SOUBOR

Data byla získaná v rámci tréninkové matice z předmětu Základy statistiky v sociálních vědách. Známé jsou proměnné jako pohlaví, věk a třída a informace o tom, jestli cestující přežil či nikoliv. Cílem analýzy je zjistit, které proměnné mají statisticky významný vliv na pravděpodobnost přežití pasažérů na Titaniku. Jsem si vědoma, že taková data patří mezi nejčastěji používané sady dat pro výuku statistických metod. I přesto, že lze odhadnout, jaký výsledek analýza bude mít, jsem se rozhodla vyzkoušet data zanalyzovat sama. Data jsou přehledná a vhodná pro vyzkoušení aplikace binární logistické regrese. Analýza byla provedena ve statistickém programu SPSS. Data obsahují informace o 1313 cestujících na palubě Titanic.

Data a další informace o této zprávě jsou dostupné na adrese: <https://dostal.vyzkum-psychologie.cz/stat4?i=471>

ZÁVISLE PROMĚNNÁ

- **SURVIVED** (dichotomická proměnná)
 - Kódování: 0 = nepřežil; 1 = přežil

NEZÁVISLE PROMĚNNÁ (PREDIKTOR)

- **PCLASS** (Ticket class)
- **SEX** (Sex)
- **AGE** (Age in years)

VÝSLEDKY

TABULKA 1 (VARIABLES IN THE EQUATION)

<u>Variables in the Equation</u>						
		B	S.E.	Wald	df	Sig.
Step 1 ^a	<u>PClass</u>			83,292	2	<,001
	<u>PClass(1)</u>	2,531	,278	82,847	1	<,001
	<u>PClass(2)</u>	1,206	,238	25,666	1	<,001
	<u>Age</u>	-,041	,008	28,565	1	<,001
	<u>Sex</u>	2,665	,203	172,501	1	<,001
	<u>Constant</u>	-1,324	,256	26,679	1	<,001

a. Variable(s) entered on step 1: PClass, Age, Sex.

- **B** = koeficient regrese – jaký je vliv proměnné na log-šanci přežití
- **S.E.** = standardní chyba odhadu B
- **Wald** = Test významnosti pro daný koeficient
- **df** = Stupeň volnosti (většinou 1)
- **Sig.** = P-hodnota (statistická významnost)
- **Exp(B)** = Odds ratio – o kolik se změní šance na přežití při změně dané proměnné o 1 jednotku

PClass: B = 2,531, Exp(B) = 12,566. První třída oproti třetí třídě: Lidé v první třídě měli cca 12,5x větší šanci na přežití než lidé ze třetí třídě. Druhá třída oproti třetí třídě: Exp(B) = 3,340. Lidé ve druhé třídě měli zhruba 3,3 x vyšší šanci na přežití než ti ve třetí třídě. Výsledek je statisticky významný ($p < 0,001$).

Age: B = -0,041, Exp(B) = 0,960. Se stoupajícím věkem se šance na přežití snižuje. Každý rok věku snižuje šanci na přežití. Výsledek je statisticky významný ($p < 0,001$).

Sex: B = 2,665, Exp(B) = 14,364. Ženy měly cca 14x vyšší šanci na přežití než muži.

Data a další informace o této zprávě jsou dostupné na adrese: <https://dostal.vyzkum-psychologie.cz/stat4?i=471>

Na základě výsledků binární logistické regrese lze konstatovat, že všechny zvolené prediktory, tedy cestovní třída (PClass), věk (Age) a pohlaví (Sex) jsou statisticky významně ovlivňují pravděpodobnost přežití cestujících (všechny proměnné mají p-hodnotu < 0,001).

Nejsilnějším prediktorem přežití bylo pohlaví. Ženy měly přibližně 14,4× vyšší šanci na přežití než muži. Významný vliv měla také cestovní třída – pasažéři v první třídě měli zhruba 12,6× vyšší šanci na přežití než cestující ve třetí třídě. Cestující ve druhé třídě měli 3,3× vyšší šanci než ti ve třetí třídě.

Věk měl negativní vliv na pravděpodobnost přežití. S každým dalším rokem života se šance na přežití snížila přibližně o 4 %, což může souviset s vyšší zranitelností starších cestujících.

Výsledky modelu ukazují, že přežití na Titaniku významně ovlivňovaly sociální faktory (třída, pohlaví) i biologický faktor (věk).

TABULKA 2 (CLASSIFICATION TABLE)

		Predicted		Percentage Correct
Observed		Survived 0	1	
Step 1	Survived 0	370	71	83,9
	1	91	222	70,9
	Overall Percentage			78,5

a. The cut value is ,500

- „Observed“ (skutečné hodnoty) – co opravdu bylo (přežil / nepřežil),
- „Predicted“ (modelové předpovědi) – co předpověděl model,
- Percentage Correct – jak přesný model byl.

Model binární logistické regrese klasifikoval přežití s celkovou přesností 78,5 %, přičemž správně identifikoval 83,9 % nepřeživších a 70,9 % přeživších.

TABULKA 3 (OMNIBUS TESTS OF MODEL COEFFICIENTS)

		Omnibus Tests of Model Coefficients		
		Chi-square	df	Sig.
Step 1	Step	334,231	4	<,001
	Block	334,231	4	<,001
	Model	334,231	4	<,001

Tabulka ukazuje, jestli celý model jako celek (tedy všechny prediktory dohromady) statisticky významně zlepšuje predikci oproti modelu bez prediktorů (nulový model).

Chi-square = 334,231, df = 4, $p < 0,001$: Model je statisticky významný. Zahrnuté proměnné (PClass, Age, Sex) jako celek významně přispívají k vysvětlení pravděpodobnosti přežití.

ZDROJE:

- Encyclopaedia Britannica. (n.d.). Titanic. <https://www.britannica.com/topic/Titanic>
- Masarykova univerzita. (2017). Binomická logistická regrese [Prezentace]. In Multivarianní analýza. <https://is.muni.cz/el/sci/jaro2017/MAS02/um/67880443/Prednaska12.pdf>